



EDITORIAL

Spanish influenza score: poder predictivo sin renunciar a lo clásico

Spanish influenza score: Predictive power without giving up the classic

J.L. García Garmendia

Unidad de Cuidados Intensivos, Servicio de Cuidados Críticos y Urgencias, Hospital San Juan de Dios del Aljarafe, Bormujos, Sevilla, España

Disponible en Internet el 17 de octubre de 2020

En el presente número de Medicina Intensiva publicamos un trabajo realizado sobre un registro multicéntrico español de gripe grave, que desarrolla una escala pronóstica de mortalidad en las UCI¹.

En los años 80 del siglo XX, se inició para los cuidados intensivos un camino de comprensión de la realidad y de ayuda a la toma de decisiones basado en escalas de gravedad. El APACHE II, aún con matizaciones, sigue vigente en la valoración de la gravedad de los pacientes críticos. Los mecanismos de generación de estas escalas se basaban en la acumulación de gran cantidad de datos representativos, y la utilización de técnicas de análisis multivariante a través de regresión logística (RL) que permitieran generar modelos predictivos y utilizar los estimadores beta para generar las puntuaciones individuales. Para los investigadores y los clínicos, el nivel de comprensión de los entresijos matemáticos necesarios para la obtención de los estimadores beta es suficiente, y existe una razonable correlación de entendimiento con las *odds ratio* y sus intervalos de confianza, formando un conjunto metodológico comprensible e interpretable².

Por otra parte, el uso de técnicas de análisis masivo de datos a través del *machine learning* (ML) ha tenido un crecimiento exponencial en los últimos años, como puede comprobarse en el número de citas bibliográficas registradas en Medline³. Sin embargo, uno de los problemas de las técnicas de ML reside en la dificultad para trasladar los análisis

al ámbito de la práctica clínica⁴. A diferencia de las técnicas convencionales de análisis estadístico, los resultados de los estudios publicados tienen buenos indicadores matemáticos, pero la percepción de los clínicos es de una aplicabilidad práctica limitada⁵. En parte, esto es debido a la dificultad para comprender los mecanismos a través de los cuales se generan los resultados y a las dificultades de utilizar un gran número de variables de manera simultánea. Ello probablemente es más concordante con la compleja realidad biológica, pero reduce la posibilidad de manejarlo en entornos de práctica clínica para los profesionales.

En este sentido, el trabajo que presentamos en este número de Medicina Intensiva, consigue un doble objetivo: incorporar técnicas de aprendizaje automático a una extensa base de datos de gripe grave en las UCI, y generar una escala de riesgo de mortalidad combinando este abordaje con otras técnicas clásicas, más asequibles para su incorporación a la práctica clínica.

La gripe grave constituye cada invierno un reto para las UCI de todo el mundo. Con el brote de gripe A (H1N1) del año 2009 como uno de los más importantes, han sido varias las temporadas que han generado problemas asistenciales en las UCI, afectando a pacientes jóvenes, generando distrés respiratorio severo, con estancias prolongadas y una elevada mortalidad⁶.

La comparación de los resultados obtenidos por técnicas convencionales con las obtenidas mediante un análisis avanzado con *random forest* (ML) da fortaleza a los hallazgos, y parece indicar que las nuevas técnicas serán capaces de añadir información a las técnicas de análisis clásico, pero

Correo electrónico: joseluis.garciagarmendia@sjd.es



gran parte de la información sustantiva puede lograrse con estas últimas⁷. Eso sí, para que las técnicas de RL tengan consistencia debemos tener registros de calidad y de tamaño muestral suficiente, lo que se ha garantizado en este estudio, a diferencia de otros recientes en los que un tamaño muestral insuficiente potencia la capacidad predictiva del ML sobre la RL⁸.

El desarrollo de una escala de predicción de mortalidad en pacientes críticos con gripe grave puede servir de ayuda para la toma de decisiones de ingreso, la toma de decisiones terapéuticas (prono, oxigenación extracorpórea, óxido nítrico) o incluso las decisiones de traslado para aplicar técnicas avanzadas en otros centros. Otra utilidad de esta escala reside en la posibilidad de segmentar grupos de riesgo para orientar ensayos terapéuticos, así como para el *benchmarking* de unidades. La utilización de variables presentes en el momento del ingreso en este estudio debe considerarse además una ventaja, pues facilitaría la asesoría precoz en la toma de decisiones. Algunos modelos que utilizan variables de curso clínico pueden valer para comparar resultados de diferentes unidades, pero no para establecer pronósticos precoces en las primeras horas de ingreso o para definir grupos susceptibles de ensayos terapéuticos.

El estudio, no obstante, tiene algunas limitaciones. La base de datos es amplia y multicéntrica, pero abarca un periodo muy amplio de tiempo (10 años), en el que se han modificado las estrategias terapéuticas y los resultados. Aunque se realiza una validación interna segmentando la base de datos, es preciso evaluar la utilidad de la escala de manera prospectiva, para corroborar la exactitud de las predicciones. Por otra parte, analiza la mortalidad en las UCI, y la escala APACHE II está diseñada para la mortalidad hospitalaria, mientras que la escala SOFA no fue ni diseñada con ese fin. Tampoco podemos descartar que el uso de ML con más variables registradas hubiera tenido mayor poder predictivo.

El futuro de las técnicas de análisis a través del ML se dirigirá casi seguro al asesoramiento en tiempo real de la actividad clínica con retroalimentación inmediata y enriquecimiento de los procesos analíticos⁹. Aunque este escenario llegará, habrá que valorar el poder que tendrá esa información en la toma de decisiones, desde un punto de vista ético, legal y deontológico¹⁰. Y tendrá que dilucidarse el papel del clínico en la aplicación y en la retirada de tratamientos cuando el propio sistema de aprendizaje automático se alimente de las decisiones que él llegue a inducir. Serán problemas de las nuevas generaciones, y la casi imposibilidad de comprender el modo de funcionamiento matemático generará complejas sensaciones en los profesionales. Mientras eso ocurre, habrá que seguir apostando por el desarrollo de herramientas asequibles y válidas como la presentada en este número.

La medicina intensiva trabaja localmente con pocos pacientes, y si ha de centrarse en enfermedades concretas, las limitaciones son aún mayores. De aquí la importancia de

disponer de potentes registros multicéntricos que faciliten análisis complejos y nos permitan añadir conocimiento en áreas de difícil manejo, y con impacto en la salud de los ciudadanos. Con la vigencia de la pandemia de la COVID-19, es una llamada a la generación de registros colaborativos.

Financiación

El autor declara no haber recibido financiación para la realización de este trabajo.

Bibliografía

1. Spanish Influenza Score (SIS). Usefulness of machine learning in the development of an early mortality prediction score in severe influenza. *Med Intensiva*. 2020; <http://dx.doi.org/10.1016/j.medin.2020.05.017>.
2. García Garmendia JL, Maroto Monserrat F. Interpretation of statistical results. *Med Intensiva*. 2018;42:370–9, <http://dx.doi.org/10.1016/j.medin.2017.12.013>.
3. Guo Y, Hao Z, Zhao S, Gong J, Yang F. Artificial Intelligence in Health Care: Bibliometric Analysis. *J Med Internet Res*. 2020;22:e18228, <http://dx.doi.org/10.2196/18228>.
4. Van Calster B, Verbakel JY, Christodoulou E, Steyerberg EW, Collins GS. Statistics versus machine learning: Definitions are interesting (but understanding, methodology, and reporting are more important). *J Clin Epidemiol*. 2019;116:137–8, <http://dx.doi.org/10.1016/j.jclinepi.2019.08.002>.
5. Núñez Reiz A, Armengol de la Hoz MA, Sánchez García M. Big Data Analysis and Machine Learning in Intensive Care Units. *Med Intensiva*. 2019;43:416–26, <http://dx.doi.org/10.1016/j.medin.2018.10.007>.
6. Sarda C, Palma P, Rello J. Severe influenza: Overview in critically ill patients. *Curr Opin Crit Care*. 2019;25:449–57, <http://dx.doi.org/10.1097/mcc.0000000000000638>.
7. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol*. 2019;110:12–22, <http://dx.doi.org/10.1016/j.jclinepi.2019.02.004>.
8. Hu CA, Chen CM, Fang YC, Liang SJ, Wang HC, Fang WF, et al. Using a machine learning approach to predict mortality in critically ill influenza patients: A cross-sectional retrospective multicentre study in Taiwan. *BMJ Open*. 2020;10:e033898, <http://dx.doi.org/10.1136/bmjopen-2019-033898>.
9. Feretzakis G, Loupelis E, Sakagianni A, Kalles D, Martsoukou M, Lada M, et al. Using Machine Learning Techniques to Aid Empirical Antibiotic Therapy Decisions in the Intensive Care Unit of a General Hospital in Greece. *Antibiotics (Basel)*. 2020;9:50, <http://dx.doi.org/10.3390/antibiotics9020050>.
10. Lazcoz Moratinos G, de Miguel Beriain I. Big Data Analysis and Machine Learning in Intensive Care Medicine: Identifying new ethical and legal challenges. *Med Intensiva*. 2020;44:319–20, <http://dx.doi.org/10.1016/j.medin.2019.11.003>.